

# 情報通信プロジェクト データ解析 -人工知能によるサッカーくじ予想-

14EC020 内田奏

14EC052 小林将輝

14EC086 中村文香

## 1 概要

サッカーくじ toto は、独立行政法人日本スポーツ振興センターにより運営・発売が行われている公営ギャンブルである。toto は国内外で行われるサッカーの試合を対象とし、購入者は勝ち、負け、引き分けを予想する。主要な対象は国内リーグの J リーグであり、対象試合には J1・J2・J3 の最大 28 試合のうち 13 試合が選ばれる。ここでは、過去の J リーグの試合結果をもとに、勝ち、負け、引き分けを予測する。

## 2 収集・使用するデータ

J リーグの試合結果はすべて J リーグ公式のデータサイト (<https://data.j-league.or.jp>) より取得可能である。取得したデータを整形し、それらを入力データとする。データ収集・整形の分担は表 1 を予定している。

表 1. データ収集・整形 分担表

役割	担当者
データ収集	内田
データ整形	小林・中村

J リーグは、toto が開始された 2000 年に初めて引き分けを導入し、1999 年以前は延長戦および PK 戦により勝敗を決していた。また、2000 年から 2002 年に導入された引き分けは、延長戦で V ゴールが記録されなかった場合のみ適用されるものであった。現在では、いかなる場合においても延長戦は行わない。このように試合方式に相違があるが、ここでは次のように勝ち、負け、引き分けを定義する。

- 1999 年以前の試合結果について
  - － 90 分及び延長での得点を勝敗として採用する。
  - － PK 戦に突入した試合は、PK 戦の結果にかかわらず引き分けとする。
- 2000 年から 2002 年の試合結果について
  - － 公式記録の勝敗を採用する。

- － V ゴールによる勝利も勝利とする。

- 2003 年以降の試合結果について

- － 公式記録の勝敗を採用する。

これによる勝ち点等公式記録との差異は今回考慮しない。

リーグ方式については、全年度 1 ステージ制であると仮定し、2 ステージ制によるリーグの動向やプレーオフ等については考慮しない。

## 3 使用するツール

学習にはニューラルネットワークを用いる。ニューラルネットワークを用いたディープラーニングフレームワークおよびツールは多くリリースされているが、今回は次のフレームワークを用いることを考えている。

- TensorFlow

時系列データの学習に長けたリカレントニューラルネットワーク (RNN) 及び LSTM についてのドキュメントや実装例が比較的豊富。

- Chainer

Python による実装が可能。

- NVIDIA DIGITS

Caffe をラップして GUI で使用可能。画像分類のための知識が筆者に多少ある。

## 4 工程

作業工程については、表 2 のスケジュールを予定している。

表 2. 工程表 (予定)

作業項目	時期
データ収集	5 月中
データ整形	6 月中
シミュレータ作成	7 月中
ネットワーク作成	8 月中
シミュレーション実行	上記が終わり次第